

Algorithm Design of Gesture Word Recognition Based on Vision

Yanling Zhang^{1, a}, Zhengqiang Yang^{2, b}, Yunfeng Wang^{3, c}

¹ School of Electronic Information Engineering, Xi'an Technological University, Xi'an, 710032, China

² Xi'an Technological University, Xi'an, 710032, China

³ Xi 'an Xu Tai Electrical Technology co., LTD ,Xi'an, 710000, China

^azylyanling@sohu.com ^b kuxm@163.com ^c214815542@qq.com

Keywords: gesture word recognition; threshold; gesture segmentation; centroid location; extreme algorithm

Abstract: Through the analysis and process of visual images of gesture words, this paper realizes the recognition of the gesture words, and this will lay a good foundation to recognize gesture language. Firstly, the acquired images are preprocessed by denoising; secondly the gesture images are segmented from the background images by setting their thresholds, and the outline of the gesture images information is obtained under morphology operations; thirdly by extracting centroid coordinates of the target gestures, the gesture images are positioned and segmented; in the end, the gesture words can be identified by calculating the extreme values to the modules. The above algorithm is validated through the Matlab software. The experimental results show that the algorithm is reliable and has good robustness.

Introduction

With the rapid development of the computer technology and virtual reality technology, multi-modal human-computer interaction technology in people's life becomes more and more important, and the detection and identification of gesture language has become an important research in the field of human-computer interaction and pattern recognition. The recognition of gesture language enables deaf people to use TV, telephone, movies, network, computers and other high-tech products, and also enables them to improve their communication with other people in a easy and fast way, greatly improving their quality of life^[1]. As an action language, gesture language can also be used between people who speak different languages. In addition, in some noisy environments, gesture language communication may be more convenient. In short, the study of gesture language recognition and realization is not only an important academic value, but also has wider social significance and a practical future.

Gesture word recognition algorithm

The algorithm is divided into image acquisition, image preprocessing, image segmentation, gesture extraction, gestures positioning, gesture image database, and gesture word recognition. On obtaining images which contain gesture words, firstly the images are preprocessed in the way of gray-scale and noise reduction. Secondly, the images are binarized and operated morphologically. Thirdly, the gesture image edges are detected trimmed; the gesture images, positioned and segmented. Finally, the gesture words can be identified by calculating the extreme values to the modules. The whole process of the algorithm is shown in Figure 1:

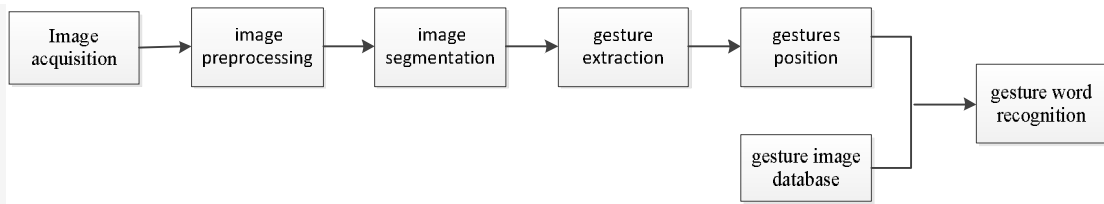


Figure 1:

Block diagram of the gesture word recognition algorithm

Image acquisition

The obtaining of gesture word images generally contains two ways: 1) through video stream, which means that the video camera, controlled by a number of functions, captures the video stream. The user can set the video rate before capturing, and each captured single frame of image is stored in the form of DJB^[2]. 2) through direct use of a camera, in which each image stands for a gesture word. In this paper, the second method is adopted.

Image preprocessing

Because the captured images are generally colorful ones, in order to reduce the complexity of gesture word recognition, we need to change the colorful ones into gray-scaled ones, that is, to change the RGB ones into gray-scaled ones. Furthermore, since there exists the impact of light, weather, and time against a variety of backgrounds, we have to denoise all these images. Meanwhile, since the median filter reduces noise in a good effect, in this case the process can be done smoothly with obvious small changes in gray values, and the vague rate around the image edges can be reduced^[3,4]. Because of all these advantages, the median filter operator is used for image filtering.

Gesture extraction

Gesture extraction is mainly used for image thresholding and morphological operations to detect the target gesture. The purpose of thresholding the image is roughly to segment the target gestures, so the threshold selection is very important, which matters whether the complex backgrounds and the target gestures can be successfully separated. In this paper we effectively obtained the threshold to segment the images by using gray-scale mean and square deviation and with Bayesian minimum error criterion as convergence conditions^[5]. Let the image $f(x, y)$ have the gray-scale range of $[a, b]$ and the binarization threshold T ($a < T < b$), the two values of the basic formula for the binarization values is:

$$f(x, y) = \begin{cases} \mathbf{1}, & \text{while } f(x, y) \geq T \\ \mathbf{0}, & \text{while } f(x, y) < T \end{cases} \quad (2.1)$$

After the threshold is calculated and the gray-scaled images are binarized by Equation (2.1), the rough target gestures can be segmented.

At this time the rough images of the target gesture, separated from their backgrounds, have been obtained. But from the obtained images, some slightly noisy points can still be seen outside the target gesture. To clear these noisy points, you need to use morphological erosion operation to get a more complete picture of the target gesture, which provides a good foundation for the further target recognition.

Gesture positioning

Gesture positioning is done to pre-match the target gesture image with that in the gesture image database. By finding the centroid position of the target gesture, the target gesture image can be cut into a size-fixed image with the centroid as the center. The edge information of the target gesture shows the most basic features of its recognition, and in order to find the centroid, we need to detect the edge of the binary image. Since the edge detected by using Prewitt edge operator has a good continuity, in this paper, the operator is adopted to detect the edge. To the breakpoints around the contour lines of the target gesture, we use morphological operations to trim and to separate the target outline from the image background only to obtain a clear gesture image.

All the pixels which Pixel has pixel value 1 will be found based on the obtained gesture

contours and their sum can be calculated. Then the sum is divided by the length of their coordinates, and the centroid of each gesture contour can be found^[6]. According to the coordinates of the centroid, the gesture image is cropped, the cropped image should be the same size as that in the gesture image database.

Gesture modeling

The aim of gesture modeling is to create a sample database including the characteristics of the various gesture contours, then the characteristics are compared between gesture image to be recognized and gesture image in the sample database only to find the most matched gesture image, and ultimately identify the word representing the gesture image. The images in the sample database come from the digital camera and the net of China Association of the Deaf. All images in the sample database are processed in the ways above and converted into a format of 1200 * 900 pixels, this is our gesture image database.

Gesture recognition

In this paper, the method of calculating the extreme values to the modules is used for the gesture words recognition. Minimum mean square deviation will be found through comparing each image pixel between the image processed to be identified and that in the gesture image database, then the corresponding sample image is the one to be recognized, so we can identify the gesture words of gesture images.

The formula of mean square deviation is as follows:

$$f(x) = [M(i, j) - Y(i, j)] \quad (2.2)$$

in which M (i, j) is the pixel of the gesture contour image to be recognized , Y (i, j) is the pixel in the gesture image database, and i, j stand for the rows and columns of images.

$$g(x) = M(i, j)^2 \quad (2.3)$$

$$S(x) = \frac{\sum_{n=1}^{i+j} f(n)}{\sum_{n=1}^{i+j} g(n)} \quad (2.4)$$

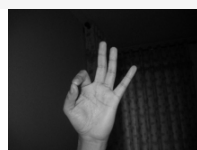
The minimum mean square deviation can be calculated from Equation (2.4), so we can identify the most matched gestures. Finally, the computer will output the most matched gesture image in the gesture image database and the corresponding prompts (text or voice), so gesture words recognition has been realized.

Experimental results and analysis

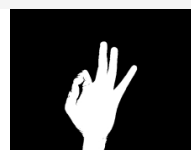
Images used in the experiment to be identified were taken by digital camera with 4000*3000 pixels color photographs, the gestures include a variety of gesture language gestures in the gesture image database. The size of binary image of gesture language for the image database is 500* 280 pixels. The hardware environment implementation of the system: PC/Celeron2.66GHz, 1G RAM; software environment: WINDOWSXP operating system, the Matlab software development platform. Part of the identification results in the experiment is as follows:



truecolor image



gray image



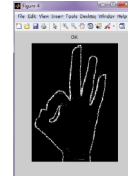
corroded Binary image



edge Trimmed image



Sheared image



recognition result

The experimental results show that the algorithm has some practical value. The algorithm has a reliable recognition to gesture words in most cases when the background is a little complicated, but it still requires further study to recognize the angle transforming gestures and gestures similar to each other.

Conclusion

Gestures extraction and centroid positioning are used to gesture language recognition in this paper, this greatly reduces the complexity of the recognition, effectively solves translation problems of the recognition, and at the same time, the speed and rate of gesture words recognition have also been improved. However, the algorithm is still much to be desired, such as the bad gesture recognition rate when the gestures are similar, etc. Generally to say, vision-based gesture language recognition has a good prospect. And it is believed that its extensively further research will positively promote the development of human-computer interaction.

References:

- [1] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, Richard Moore. Real-time human pose recognition in parts from single depth images[J]. Communications of the ACM . 2013 (1).
- [2] Thomas Bader. Video-Based Hand Gesture Recognition: Requirements and Realization for the Interaction at Tabletops. Technisches Messen . 2008
- [3] Lee C, Xu Y. Online interactive learning of gestures for human/robot interfaces. In Proceeding of IEEE Int. Conf. on Robotics and Automation. 1996, 3(1): 30~42.
- [4] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing Second Edition [M], 2004.
- [5] Bogdan Ionescu, Didier Coquin, Patrick Lambert, Vasile Buzuloiu. Dynamic Hand Gesture Recognition Using the Skeleton of the Hand. EURASIP Journal on Applied Signal Processing . 2005
- [6] J. Lee and T.L. Kunii. Model based analysis of hand posture. IEEE Computer Graphics and Applications, Sept. 1995: 77~86.